Cell, Vol. S116, S93–S96, January 23, 2004 Copyright ©2004 by Cell Press

The 20 Years It Took to Recognize the Importance of Tiny RNAs

Commentary

Gary Ruvkun,[1,*] Bruce Wightman,[1,2] and Ilho Ha[1,3]
[1]Department of Molecular Biology
Massachusetts General Hospital
Department of Genetics
Harvard Medical School
Fruit Street
Boston, Massachusetts 02114

Our interest in the *C. elegans* heterochronic genes began during Gary Ruvkun's post-PhD thesis defense seminar tour of Europe in November, 1981. He took along a Xerox of one paper, a 1981 *Cell* paper by Marty Chalfie, Bob Horvitz, and John Sulston, describing the detailed cell lineage analysis, but not the molecular identity, of two genes that affect *C. elegans* developmental timing, *lin-4* and *unc-86* (Chalfie et al., 1981). The paper was replete with specialized language and concepts that he could not decipher; his bacterial genetics training did not prepare him for the patois of *C. elegans* developmental genetics, a product of the island tribe that evolved around Sydney Brenner at the Medical Research Council labs in Cambridge, England.

Ruvkun visited the MRC on that trip and spent a few hours talking with Marty Chalfie about *lin-4* and *unc-86*. That one afternoon at the MRC planted a seed: Ruvkun glimpsed the worm community, its ambition, its exuberance, its collaborative reflexes, its sense of mission. And the field seemed ready to explode at that moment—there were lots of interesting mutants that were a few technical developments away from exciting molecular discovery. The attraction of *C. elegans* developmental genetics reasserted itself after Bob Horvitz gave a departmental seminar at Harvard in January of 1982 that was just as confusing and interesting to Ruvkun as the 1981 *Cell* paper. He went to MIT to talk about worms with Horvitz. Horvitz was very enthusiastic about cracking the problem of going molecular with these very promising genes identified by their genetics. Meeting with Horvitz was Ruvkun's second glimpse of the MRC worm culture, now transplanted to MIT by Horvitz, where it also transmuted to include a sense of urgency. Ruvkun began to work on the problem, part time in Wally Gilbert's lab, where he had begun his postdoctoral work and which was a center of molecular biology expertise, and part time in the Horvitz lab, where he would learn worm genetics.

Victor Ambros had just finished his genetic analysis of heterochronic genes (Ambros and Horvitz, 1984). The most compelling of the genes was *lin-14*, because it had both gain-of-function and loss-of-function mutant alleles with opposite developmental timing defects. Such genetic attributes define switch genes, which were considered the keys to development in the Horvitz group at that time; the virtues of such genes were recited at nearly every group meeting. Ambros and Horvitz had also discovered that *lin-4* was a probable negative regulator of *lin-14*, so there was a developmental pathway to weave molecules into. Ambros was very keen to learn the molecular identity of *lin-14* and offered to work together with Ruvkun on the molecular analysis.

The problem was that there was essentially no method to isolate a piece of DNA corresponding to a locus defined by genetics in *C. elegans* at that point. Transposons had just been detected (Liao et al., 1983) and were thought to be responsible for spontaneous mutations, but there was no well-developed protocol for going molecular. We decided to try a few strategies, some riskier than others. One very tricky and ambitious approach sought to detect DNA changes directly at the locus using a technique of Southern cross-blotting total DNA isolated from wild-type and mutant strains, but the technical demands were too much to surmount and after a year we abandoned it (Ruvkun et al., 1990). We also tried other jackpot approaches. One view was that gain-of-function mutations might be caused by transposon insertions in the same way that retrovirus insertions activate adjacent genes, so we probed every gain-of-function mutation that had been isolated in the Horvitz lab with a transposon probe, seeking a new hybridization band. Nothing. We also sought mutations in conserved pathways before such concepts were so commonplace, probing Southern blots of many mutant *C. elegans* strains with DNA from *Drosophila* Notch and bithorax complex, as these first developmental control genes were isolated. We did not detect any changes in these genes. A *C. elegans* Notch homolog did emerge from molecular analysis of the *lin-12* locus by Iva Greenwald at about the same time (Greenwald, 1985), so the idea was not entirely stupid, but it was naive to expect DNA homology and for this reason it did not work. Even the bithorax idea was pretty good, as use of more sophisticated degenerate oligonucleotide probes by Thomas Bürglin that targeted conserved protein regions did reveal many homeobox genes, including bithorax complex homologs (Bürglin et al., 1989).

More productive for the molecular identification of *lin-14* was an RFLP approach. The transposon Tc1 had spread throughout the genome of one strain of *C. elegans*. We came up with a nifty way to find the Tc1 insertion RFLPs in the genetic region closest to *lin-14*. This allowed us to bypass the identification of hundreds of RFLPs and jump right to the *lin-14* genomic region. In this way, Ambros and Ruvkun constructed recombinant strains and cloned the particular transposon insertion loci that were nearest to *lin-14*. At the same time, Alan Coulson, Bob Waterston, and John Sulston were just beginning to assemble contigs of *C. elegans* cosmid clones. As each transposon Tc1 RFLP in the *lin-14* genomic region was cloned, it could be matched with cosmids and with extended contigs that contained those cosmids to assemble a genome map of the *lin-14* region. We still did a bit of chromosome walking, but it was

*Correspondence: ruvkun@molbio.mgh.harvard.edu
[2]Present address: Biology Department, Muhlenberg College, 2400 Chew Street, Allentown, Pennsylvania 18104.
[3]Present address: BioScience Center, Hanwha Chemical R&D Institute, 6 Shingsung Dong, Yusung Gu Daejon 305-345, South Korea.

more like chromosome long jumping with the contig assembly of the MRC genome group. The coup de grace of this *lin-14* RFLP mapping was Ambros' identification of a recombinant chromosome that separated a *lin-14* gain-of-function allele from an intragenic loss-of-function mutation he had also isolated, and Ruvkun's mapping of this recombination event with other RFLPs to a few kilobases. The first evidence that *lin-14* was identified came from the detection of DNA changes in these few kilobases associated with two different *lin-14* gain-of-function alleles (Ruvkun et al., 1989).

At that point in 1985, Ruvkun moved on to a faculty position in the Department of Molecular Biology at Massachusetts General Hospital and Department of Genetics at Harvard Medical School and recruited his first students and postdocs. Ambros had moved to a faculty position just up the road at Harvard the year before. Over the next few years, Bruce Wightman, Prema Arasu, Joe Gatto, John Giusto and Thomas Bürglin determined that *lin-14* gain-of-function mutations affect the *lin-14* 3′ UTR but do not affect transcript levels, suggesting that translation of the *lin-14* mRNA is negatively regulated by *lin-4* (Ruvkun and Giusto, 1989). They found that the *lin-14* coding region was not homologous to any other protein in the then very sparse databases. To this day, *lin-14* homologs have been detected only in other nematodes, suggesting that the protein component of the *lin-4/lin-14* regulatory circuit is drifting fast or is an invention of the Nematoda. They showed that LIN-14 protein is nuclearly localized, suggesting that it may regulate gene expression. They also showed that the expression of LIN-14 protein is graded over time and that graded expression is disrupted in the *lin-4* or *lin-14* gain-of-function mutations, but mRNA levels are unaffected (Wightman et al., 1991, 1993; Arasu et al., 1991). Because the *lin-14* gain-of-function mutations mapped to the 3′ UTR and caused similar molecular defects as the *lin-4* reduction-of-function mutation, they predicted that the *lin-4* gene product would regulate the *lin-14* 3′ UTR. But they were definitely envisioning a regulatory protein that might engage the *lin-14* 3′ UTR. When the Ambros group showed that *lin-4* actually encodes an RNA, the idea of an RNA::RNA interaction emerged.

On the evening of June 11, 1992, Ambros and Ruvkun exchanged the *lin-4* and *lin-14* 3′ UTR sequences and each detected the multiple elements in the *lin-14* mRNA that are partially complementary to the *lin-4* RNA. It was a moment when years of work came together into a clear model—a classic eureka moment. One *lin-4* sequence element focused the search. We guessed that the *lin-4(ma161)* G to U point mutation would constitute an "active site" of the *lin-4* RNA and would be located in any RNA duplex with the *lin-14* 3′ UTR. It was. Other *lin-14* sequence comparisons and mutations instantly validated the *lin-4* complementary sites. We asked if the *lin-4* complementary regions are affected by *lin-14* gain-of-function mutations and are conserved in *C. briggsae*. The two *lin-14* gain-of-function mutations were neatly explained by the sites complementary to the *lin-4* RNA: the weaker gain-of-function allele deletes 5 of 7 complementary sites, whereas the stronger allele removes them all. We had shown that the *C. elegans lin-14* 3′ UTR is temporally regulated in *C. briggsae* and that there are multiple conserved sequence elements and long stretches with no conserved sequence. Candy Lee, Rhonda Feinbaum, and Ambros had also shown that the *lin-4* RNA sequence is conserved in *C. briggsae*, so we expected that *lin-14* sequence elements important for *lin-4* regulation would be conserved (Wightman et al., 1993). They were. Perfectly. In fact, the conservation of the *lin-14* mRNA sequences strongly supported the existence of distinct RNA duplex structures (e.g., bulged C versus more perfect duplexes), which we later proved to be correct (Ha et al., 1996).

To further test the model and explore the mechanism of the regulation, Wightman and Ilho Ha fused the *lin-14* 3′ UTR onto a reporter gene and showed that it is sufficient to generate graded temporal expression (Wightman et al., 1993). This showed that the *lin-4* antisense RNA does not depend on other *lin-14* mRNA sequences (the 5′ end, for example), constraining the mechanism of *lin-4/lin-14* RNA duplex regulation of translation. By monitoring LIN-14 protein and mRNA levels, Wightman also showed that the interaction does not regulate *lin-14* mRNA abundance, but rather translation of the *lin-14* mRNA. By monitoring β-galactosidase activity and *lacZ* mRNA expression from a *lacZ/lin-14* 3′ UTR fusion gene in a wild-type or the *lin-4* mutant background, Wightman and Ha also showed that *lin-4* acts posttranscriptionally via the *lin-14* 3′ UTR to generate graded temporal expression. This model was later validated and extended by Phil Olsen and Ambros, who showed that the translational control of *lin-14* occurs at a postinitiation step, because the *lin-14* mRNA that is not translated after *lin-4* expression is upregulated is paradoxically localized to polysomes (Olsen and Ambros, 1999). The localization of miRNAs to polysomes has recently been shown to be true for mammalian brain miRNAs as well, suggesting that it is general to miRNAs (Kim et al., 2003).

The combination of the Lee et al. (1993) and Wightman et al. (1993) *Cell* papers made a strong case for a direct interaction between the *lin-4* RNA and the *lin-14* mRNA. In the usual bloodbath of publishing, the reviewers wanted more, more, more, and the editors, as usual, sent along noncommittal form letters to prompt us to do more experiments. So we wrote the usual missives passionately arguing for publication. The most germane request, which we had expected, was for proof of the RNA duplex model by constructing compensatory mutations in *lin-4* and *lin-14*. In fact, the Ambros and Ruvkun labs tried to do this together but, due to instability of the two different and increasingly sophisticated engineered *lin-4* mutants, we could not do the elegant experiment. We later did show that the *lin-4* RNA binds to the *lin-14* mRNA in vitro and that a *lin-14* 3′ UTR with mutations in each of the *lin-4* binding sites is no longer downregulated by *lin-4* in vivo nor bound in vitro (Ha et al., 1996). But because there were many experimental supports of the model besides allele-specific suppression, and because the strategy failed a few times, we abandoned elegance. Ruvkun continues to argue that elegance in molecular genetics is aesthetically pleasing but scientifically overrated.

After the two papers were published, Marv Wickens published a very nice News and Views in *Nature* that ended with the conjecture that tiny RNAs and their tar-

gets might be more extensive than just *lin-4* and *lin-14* (Wickens and Takayama, 1994). But the discovery of the world's first microRNA did not trigger a gold rush, not even by the Ambros or Ruvkun labs. First, the heterochronic pathway was a rather parochial object of study; while the *lin-4* and *lin-14* stories were published in high-profile journals, they were viewed as a novelty rather than a harbinger. The antisense regulation was similar to some prokaryotic gene regulatory vignettes, if one ignored how incredibly small *lin-4* was (and *lin-4* was four times smaller than any other noncoding regulatory RNA). Without homologs in other species, its generality did not emerge. We still did not suspect an extensive microRNA world even after Brenda Reinhart, Frank Slack, and others detected a second microRNA, *let-7* (Reinhart et al., 2000), because it emerged from genetic analysis of the same *C. elegans* heterochronic pathway; tiny RNAs could still have been inventions of this one pathway in this one species.

We began to suspect an extensive tiny RNA world after genome database searches using the *let-7* miRNA sequence revealed perfect 22 nt matches in the newly emerging *Drosophila* and human genome sequence. The genome regions adjacent to the perfect matches could also fold into bulged and loopy precursors that looked a lot like the probable *lin-4* and *let-7* precursors of *C. elegans*, suggesting that these 22 nt perfect matches were not spurious. It is important to stress that the detection of these homologs demanded full genome databases, not the biased "protein world" information of EST databases. Amy Pasquinelli, Brenda Reinhart, and Ruvkun confirmed that the fly and human *let-7* homologs detected in databases express a 22 nt RNA, and in collaboration with a large number of people who sent RNA samples, they showed that *let-7* is conserved across most of animal phylogeny, analyzing RNAs from a very satisfying range of nondomesticated animals such as coral, mollusks, annelids, acorn worms, you name it (Pasquinelli et al., 2000). Even more surprising were the findings by Pasquinelli and Reinhart that the temporal regulation of *let-7* is also conserved in a wide range of species and by Slack and coworkers that complementary sites in the target of *let-7*, the *lin-41* mRNA, are also conserved across phylogeny, strongly suggesting an ancient function in temporal patterning (Pasquinelli et al., 2000; Slack et al., 2000).

The conservation of the *let-7* RNA was the key finding that argued the generality of miRNAs for our group. It was then that we began a collaboration with the Church lab to search for more microRNA genes by informatics (Grad et al., 2003). Biochemical searches for miRNAs by the Ambros, Bartel, and Tuschl labs identified some of this mother lode before we did (Lau et al., 2001; Lagos-Quintana et al., 2001; Lee and Ambros, 2001). Now we know that there are hundreds, perhaps thousands, of miRNA genes in various genomes, about a third of which are conserved. A few of these genes have now emerged from genetic analysis in *Arabidopsis* and *Drosophila*, but many more have unknown functions (Ruvkun, 2001). And while the paradigm from *lin-4/lin-14* remains the model, miRNAs have now been shown to control mRNA abundance in plants, and they could regulate many more RNA steps than translation. In addition, the assignment of the related siRNAs to chromatin silencing in *S. pombe* suggests that miRNAs could act beyond the control of mRNA abundance or translation (Volpe et al., 2002).

An even deeper connection to RNAi started with numerological considerations (it cannot be called reasoning). When siRNAs of 22 nt, the same size as *lin-4* and *let-7*, were discovered by the Baulcombe and Tuschl groups in 1999 and 2001 (Hamilton and Baulcombe, 1999; Elbashir et al., 2001), Ruvkun noted that the number 22 (the number of letters in the Hebrew alphabet) is stressed in the Kabbalah, a Jewish mystical tradition celebrated in medieval Spain, alternative bookstores, and a number of helpful Web sites (e.g., http://pws.prserv.net/leon/Kabbalah-articles/treeof.html). We began to explore the action of the RNAi machinery in miRNA maturation and activity. Amy Pasquinelli looked closely at the first RNAi-defective mutants, *rde-1* and *rde-4*, but could not detect any heterochronic defects nor any change in *lin-4* or *let-7* miRNA activity or processing. At that point, Alla Grishok and Craig Mello contacted us after they discovered that RNAi inactivation of one of 28 different RDE-1 paralogs causes a phenotype similar to the *let-7* lethality. Grishok and Pasquinelli then showed that RNAi inactivation of a pair of RDE-1 paralogs and *C. elegans* Dicer disrupt miRNA processing and activity, proving that the RNAi and miRNA pathways are related (Grishok et al., 2001). The intersection with RNAi dramatically increased the interest in miRNAs; with so many labs using RNAi as a tool, the mechanism by which miRNAs and siRNAs inhibit gene function has a large audience. Still emerging from genetic and functional genomic analysis are the components of the Dicer and RISC complexes that process dsRNA and miRNA precursors and present them to mRNAs, as well as components that recognize miRNA::mRNA duplexes to downregulate translation on polyribosomes.

It is now clear an extensive miRNA world was flying almost unseen by our genetic radar. As much as geneticists like to think that nothing can escape genetic analysis, the miRNA genes are so small that they almost escaped our notice. Now, more miRNA genes are emerging from loss-of-function genetics (Johnston and Hobert, 2003) as well as from gain-of-function genetics focused on mutations in the target genes that are negatively regulated by miRNAs (Llave et al., 2002) or based on misexpression of the miRNAs (Brennecke et al., 2003). The families of miRNAs that have emerged from informatic and biochemical analyses suggest much gene duplication and divergence, as has been seen in other large families of regulatory genes in multicellular organisms, such as transcription factor genes. The 22 nt length is just about right for specificity to particular target genes as well as for plasticity to evolve towards new gene targets. The number of protein and large RNA coding genes in multicellular animals and plants is surprisingly small, only an order of magnitude more than in microbial genomes. The flowering of the diverse and numerous miRNA genes in animals and plants may turn out to mediate much of the gene regulation that generates cell diversity and developmental patterning, as well as the gene regulation underlying other recent inventions in animals such as synaptic signaling and its modulation.

References

Ambros, V., and Horvitz, H.R. (1984). Heterochronic mutants of the nematode *Caenorhabditis elegans*. Science *226*, 409–416.

Arasu, P., Wightman, B., and Ruvkun, G. (1991). Temporal regulation of lin-14 by the antagonistic action of two other heterochronic genes, lin-4 and lin-28. Genes Dev. *5*, 1825–1833.

Brennecke, J., Hipfner, D.R., Stark, A., Russell, R.B., and Cohen, S.M. (2003). Bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in *Drosophila*. Cell *113*, 25–36.

Bürglin, T.R., Finney, M., Coulson, A., and Ruvkun, G. (1989). *C. elegans* has scores of homeobox genes. Nature *341*, 239–243.

Chalfie, M., Horvitz, H.R., and Sulston, J.E. (1981). Mutations that lead to reiterations in the cell lineages of *C. elegans*. Cell *24*, 59–69.

Elbashir, S.M., Lendeckel, W., and Tuschl, T. (2001). RNA interference is mediated by 21- and 22-nucleotide RNAs. Genes Dev. *15*, 188–200.

Grad, Y., Aach, J., Hayes, G.D., Reinhart, B., Church, G.M., Ruvkun, G., and Kim, J. (2003). Computational and experimental identification of *C. elegans* microRNAs. Mol. Cell *11*, 1253–1263.

Greenwald, I. (1985). lin-12, a nematode homeotic gene, is homologous to a set of mammalian proteins that includes epidermal growth factor. Cell *43*, 583–590.

Grishok, A., Pasquinelli, A.E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D.L., Fire, A., Ruvkun, G., and Mello, C.C. (2001). Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *Caenorhabditis elegans* developmental timing. Cell *106*, 23–34.

Ha, I., Wightman, B., and Ruvkun, G. (1996). A bulged lin-4/lin-14 RNA duplex is sufficient for *Caenorhabditis elegans* lin-14 temporal gradient formation. Genes Dev. *10*, 3041–3050.

Hamilton, A.J., and Baulcombe, D.C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. Science *286*, 950–952.

Johnston, R.J., and Hobert, O. (2003). A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*. Nature *426*, 845–849.

Kim, J., Krichevsky, A., Grad, Y., Hayes, G.D., Kosik, K.S., Church, G.M., and Ruvkun, G. (2003). Identification of many microRNAs that copurify with polyribosomes in mammalian neurons. Proc. Natl. Acad. Sci. USA, *101*, 360–365. Published online December 22, 2003. 10.1073/pnas.2333854100.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. Science *294*, 853–858.

Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001). An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. Science *294*, 858–862.

Lee, R.C., and Ambros, V. (2001). An extensive class of small RNAs in *Caenorhabditis elegans*. Science *294*, 862–864.

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843–854.

Liao, L.W., Rosenzweig, B., and Hirsh, D. (1983). Analysis of a transposable element in *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. USA *80*, 3585–3589.

Llave, C., Xie, Z., Kasschau, K.D., and Carrington, J.C. (2002). Cleavage of Scarecrow-like mRNA targets directed by a class of *Arabidopsis* miRNA. Science *297*, 2053–2056.

Olsen, P.H., and Ambros, V. (1999). The lin-4 regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. Dev. Biol. *216*, 671–680.

Pasquinelli, A., Reinhart, B., Slack, F., Maller, B., Kuroda, M., Martindale, M., Srinivasan, A., Fishman, M., Hayward, D., Ball, E., et al. (2000). Conservation across animal phylogeny of the sequence and temporal expression of the 21 nucleotide let-7 heterochronic regulatory RNA. Nature *408*, 86–89.

Reinhart, B.J., Slack, F.A., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.C., Horvitz, H.R., and Ruvkun, G. (2000). The 21 nucleotide let-7 RNA regulates *C. elegans* developmental timing. Nature *403*, 901–906.

Ruvkun, G. (2001). Glimpses of the tiny RNA world. Science *294*, 797–799.

Ruvkun, G., and Giusto, J. (1989). The *Caenorhabditis elegans* heterochronic gene lin-14 encodes a nuclear protein that forms a temporal switch during development. Nature *338*, 313–319.

Ruvkun, G., Ambros, V., Coulson, A., Waterston, R., Sulston, J., and Horvitz, H.R. (1989). Molecular genetics of the *Caenorhabditis elegans* heterochronic gene lin-14. Genetics *121*, 501–516.

Ruvkun, G., Gilbert, W., and Horvitz, H.R. (1990). Detection of mutations and DNA polymorphisms using whole genome Southern Cross hybridization. Nucleic Acids Res. *18*, 809–815.

Slack, F.J., Basson, M., Liu, Z., Ambros, V., Horvitz, H.R., and Ruvkun, G. (2000). The *lin-41* RBCC gene acts in the *C. elegans* heterochronic pathway between the *let-7* regulatory RNA and the *lin-29* transcription factor. Mol. Cell *5*, 659–669.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I., and Martienssen, R.A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science *297*, 1833–1837.

Wickens, M., and Takayama, K. (1994). RNA. Deviants—or emissaries. Nature *367*, 17–18.

Wightman, B., Bürglin, T.R., Gatto, J., Arasu, P., and Ruvkun, G. (1991). Negative regulatory sequences in the lin-14 3′ untranslated region are necessary to generate a temporal switch during *C. elegans* development. Genes Dev. *5*, 1813–1824.

Wightman, B., Ha, I., and Ruvkun, G. (1993). Post-transcriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. Cell *75*, 855–862.

Cell, Vol. 75, 843–854, December 3, 1993, Copyright ©1993 by Cell Press

The C. elegans Heterochronic Gene *lin-4* Encodes Small RNAs with Antisense Complementarity to *lin-14*

Rosalind C. Lee,*† Rhonda L. Feinbaum,*‡ and Victor Ambros†
Harvard University
Department of Cellular and Developmental Biology, Cambridge, Massachusetts 02138

Summary

***lin-4* is essential for the normal temporal control of diverse postembryonic developmental events in C. elegans. *lin-4* acts by negatively regulating the level of LIN-14 protein, creating a temporal decrease in LIN-14 protein starting in the first larval stage (L1). We have cloned the C. elegans *lin-4* locus by chromosomal walking and transformation rescue. We used the C. elegans clone to isolate the gene from three other Caenorhabditis species; all four Caenorhabditis clones functionally rescue the *lin-4* null allele of C. elegans. Comparison of the *lin-4* genomic sequence from these four species and site-directed mutagenesis of potential open reading frames indicated that *lin-4* does not encode a protein. Two small *lin-4* transcripts of approximately 22 and 61 nt were identified in C. elegans and found to contain sequences complementary to a repeated sequence element in the 3′ untranslated region (UTR) of *lin-14* mRNA, suggesting that *lin-4* regulates *lin-14* translation via an antisense RNA-RNA interaction.**

Introduction

A genetic pathway of heterochronic genes in Caenorhabditis elegans acts to specify the temporal fates of cells during larval development, thereby controlling the timing and sequence of events in diverse postembryonic cell lineages (Ambros and Horvitz, 1984; Ambros and Horvitz, 1987; Ambros, 1989). Mutations in the heterochronic genes can cause either precocious development, in which normally late developmental programs are expressed at early larval stages, or retarded development, in which normally early developmental programs are reiterated at later stages (Chalfie et at., 1981; Ambros and Horvitz, 1984). It is likely that the expression of stage-specific developmental programs by particular cells requires the accurate temporal regulation of the products of the heterochronic genes.

lin-4 acts early in C. elegans larval development to affect the timing of developmental events at essentially all larval stages and in diverse cell types (Chalfie et al., 1981; Ambros and Horvitz, 1987). Animals carrying a *lin-4* loss-of-function *(lf)* mutation, *lin-4(e912)*, display reiterations of early fates at inappropriately late developmental stages; cell lineage patterns normally specific for the L1 are reiterated at later stages, and the animals execute extra larval molts (Chalfie et al., 1981). The consequences of these heterochronic developmental patterns include the absence of adult structures (such as adult cuticle and the vulva) and the prevention of egg laying.

lin-14 null *(0)* mutations cause a phenotype opposite to that of *lin-4(lf)* and are completely epistatic to *lin-4(lf)*, which is consistent with *lin-4* acting as a negative regulator of *lin-14* (Ambros and Horvitz, 1987; Ambros, 1989). *lin-14(0)* mutants skip the expression of L1-specific events and precociously execute programs normally specific for the L2, L3, L4, and adult stages. *lin-14* gain-of-function *(gf)* mutations, which cause inappropriately high *lin-14* activity at late stages of development, result in a retarded phenotype virtually identical to that of *lin-4(lf)* (Ambros and Horvitz, 1987). These observations indicate that in wild-type development a high level of *lin-14* activity in the early L1 stage specifies L1-specific programs, and lower levels of *lin-14* activity in the late L1 specify later stage-specific programs. Thus, the normal developmental progression from the execution of L1 programs to later programs depends critically on the *lin-4*-dependent decrease in *lin-14* activity.

The temporal decrease in *lin-14* activity reflects a decrease in the level of LIN-14 protein. LIN-14 protein is normally abundant in the nuclei of late-stage embryos and younger L1 larvae and then is barely detectable by the L2 (Ruvkun and Giusto, 1989). *lin-14* transcripts are constant throughout development, indicating that *lin-14* is negatively regulated posttranscriptionally (Wightman et al., 1993 [this issue of *Cell*]). In *lin-4* mutant animals, as in *lin-14(gf)* mutants, the level of LIN-14 protein remains abnormally high late in development (Arasu et al., 1991). Mapping of the *lin-14(gf)* mutations to the 3′UTR of *lin-14* mRNA (Wightman et al., 1991), and gene fusion experiments (Wightman et al., 1993), define the *lin-14* 3′UTR as a necessary and sufficient cis-negative regulatory element of LIN-14 protein level. The temporal decrease in LIN-14 protein levels requires both *lin-4* in trans and *lin-14* 3′UTR sequences in cis. These observations suggest that the *lin-4* gene product or products act via the *lin-14* 3′UTR to inhibit, directly or indirectly, the translation of *lin-14* mRNA. To determine the mechanism by which *lin-4* developmentally regulates the level of LIN-14 protein, we cloned the *lin-4* locus by chromosomal walking and transformation rescue. Our analysis of the *lin-4* genomic sequence indicates that *lin-4* does not encode a protein product. We have identified two small *lin-4* transcripts of approximately 22 and 61 nt. These *lin-4* RNAs are complementary to a repeated sequence in the 3′UTR of *lin-14* (Wightman et al., 1991, 1993), supporting a model in which these *lin-4* RNAs could regulate *lin-14* translation by an antisense mechanism.

*The first two authors contributed equally to this work.
†Present address: Dartmouth College Department of Biology, Hanover, New Hampshire 03755.
‡Present address: Massachusetts General Hospital Department of Molecular Biology, Boston, Massachusetts 02114.

Cell, Vol. 75, 855–862, December 3, 1993, Copyright ©1993 by Cell Press

Posttranscriptional Regulation of the Heterochronic Gene *lin-14* by *lin-4* Mediates Temporal Pattern Formation in C. elegans

Bruce Wightman,*† Ilho Ha,* and Gary Ruvkun
Department of Molecular Biology
Massachusetts General Hospital
Boston, Massachusetts 02114

Summary

During C. elegans development, the temporal pattern of many cell lineages is specified by graded activity of the heterochronic gene *Lin-14*. Here we demonstrate that a temporal gradient in Lin-14 protein is generated posttranscriptionally by multiple elements in the *lin-14* 3′UTR that are regulated by the heterochronic gene *Lin-4*. The *lin-14* 3′UTR is both necessary and sufficient to confer *lin-4*–mediated posttranscriptional temporal regulation. The function of the *lin-14* 3′UTR is conserved between C. elegans and C. briggsae. Among the conserved sequences are seven elements that are each complementary to the *lin-4* RNAs. A reporter gene bearing three of these elements shows partial temporal gradient activity. These data suggest a molecular mechanism for Lin-14p temporal gradient formation: the *lin-4* RNAs base pair to sites in the *lin-14* 3′UTR to form multiple RNA duplexes that down-regulate *lin-14* translation.

Introduction

The heterochronic genes of Caenorhabditis elegans form a regulatory hierarchy that coordinately controls the temporal identities of cells during development (Ambros, 1989). Like the Drosophila anterior–posterior and dorsal–ventral axis determination genes that control cell identities specifically in their respective axes (Nüsslein-Volhard, 1991), heterochronic genes point to a third orthogonal axis under explicit genetic control, that of time. The heterochronic gene pathway generates a temporal gradient of Lin-14 protein (Lin-14p) that specifies the normal temporal sequence of cell lineages (Ruvkun and Giusto, 1989). Mutations in the heterochronic genes *lin-4*, *lin-14*, *lin-28*, and *lin-29* cause temporal transformations of cell lineage in many tissues and cell types (Chalfie et al., 1981; Ambros and Horvitz, 1984). Depending on the mutation, stage-specific cell lineages in heterochronic mutant animals occur at either earlier or later stages than they would normally. These mutations perturb temporal patterning by disrupting the regulation of, or response to, the Lin-14p temporal gradient. A key question in temporal pattern formation is how the heterochronic gene pathway generates this molecular gradient.

The instructive role of *lin-14* in temporal pattern formation was established by the observation that *lin-14* gain-of-function (gf) and loss-of-function (lf) mutations have opposite phenotypes (Ambros and Horvitz, 1987). *lin-14(lf)* alleles cause larvae stage 2 (L2) patterns of cell lineage in a variety of tissues to be executed precociously during the L1 stage (Ambros and Horvitz, 1987). Two *lin-14(gf)* alleles cause the opposite transformation in temporal cell fate, reiterations of early cell fates at later stages. For instance, at the L2 stage, *lin-14(gf)* mutants repeat patterns of cell lineage appropriate for the L1 stage (Ambros and Horvitz, 1984).

lin-14 controls these stage-specific cell lineages by generating a temporal gradient of Lin-14 nuclear protein (Lin-14p). Lin-14p levels are high during the L1 stage but are faintly detectable by indirect immunofluorescence at L2 and later stages (Ruvkun and Giusto, 1989). The *lin-14(gf)* mutations interfere with *lin-14* temporal down-regulation, causing inappropriately high Lin-14p levels at post-L1 stages, which imposes L1 patterns of cell lineage on post-L1 blast cells (Ruvkun and Giusto, 1989). These mutations delete sequences from the *lin-14* mRNA 3′ untranslated region (3′UTR), suggesting that regulatory elements in the 3′UTR normally are responsible for mediating the down-regulation of Lin-14p abundance at later larval stages (Wightman et al., 1991).

The heterochronic gene *lin-4* is a negative regulator of *lin-14* (Ambros, 1989; Arasu et al., 1991). A null mutation in *lin-4* results in L1-specific cell lineage reiterations similar to those observed with *lin-14(gf)* mutations and causes Lin-14p to be present at post-L1 stages (Chalfie et al., 1981; Arasu et al., 1991; Lee et al., 1993 [this issue of *Cell*]). In addition, *lin-14* gene activity is necessary for the Lin-4 heterochronic defects (Ambros, 1989). Therefore, the *lin-4* gene product is required for normal temporal regulation of Lin-14p and is a candidate for directly interacting with negative regulatory elements in the *lin-14* 3′UTR identified by *lin-14(gf)* mutations. The functional products of the *lin-4* gene are two small untranslated RNA molecules (Lee et al., 1993).

Here we examine the molecular mechanism by which the Lin-14p temporal gradient is generated. We find that temporal regulation of Lin-14p abundance during wild-type development occurs at a posttranscriptional step and is mediated by the *lin-14* 3′UTR that bears multiple conserved elements complementary to the *lin-4* RNAs. These data suggest that the Lin-14p temporal gradient is generated by *lin-4* antisense base pairing to the *lin-14* mRNA.

Results

Down-Regulation of Lin-14p Occurs at a Posttranscriptional Step

To obtain a quantitative description of the temporal regulation of Lin-14p previously observed in indirect immunofluorescence experiments (Ruvkun and Giusto, 1989; Arasu et al., 1991), immunoblots were performed using protein extracts from synchronized wild-type, *lin-14(gf)*, and *lin-4* mutant animals. Early stage preparations (71%–89%, L1 stage) and late stage preparations (78%–94%, L2 and L3 stages) were prepared from wild-type and mutant strains.

*The first two authors contributed equally to this work.
†Present address: Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, California 94720.